

DOI: 10.3724/SP.J.1005.2010.00473

汉族人群中 22 个 *HLA-Cw* 等位基因全长序列单核苷酸多态性分析

曾健强, 徐筠婷, 王大明, 高素青, 邹红岩, 邓志辉

深圳市血液中心, 深圳市组织配型与免疫遗传重点实验室, 深圳 518035

摘要: 为探讨 *HLA-Cw* (Human leukocyte antigen-Cw) 基因全长序列分子遗传多态性, 文章对 28 个 *HLA-Cw* 基因型已知的汉族个体样本, 采用长距离 PCR 技术和高保真性的 *Pfu* 酶, 扩增 *HLA-Cw* 基因全长序列 4.5 kb, 进行分子克隆和单倍体测序。采用群体遗传学研究方法分析了 *HLA-Cw* 等位基因全长序列中各亚区的单核苷酸多态性。结果表明: 在 28 个样本中共检测出 22 种等位基因, 序列均已提交 GenBank 和国际 IMGT/HLA 数据库并获得了认可; 其中 *Cw*0706*、*Cw*030301*、*Cw*140201* 的全长序列为首次报道, 尤其是 *Cw*0706* 内含子序列的获得, 能够重新设计对该等位基因测序分型的引物, 避免测序分型中可能对这一等位基因的漏检。将 28 个样本的 56 条单倍体序列用 Clustal 软件进行序列排比, 输入到 DnaSP4.0 进行多态性分析, 共发现 244 个 SNPs, 10 处插入/缺失多态性。对 *HLA-Cw* 等位基因各亚区多态性的分析, 发现第 4 内含子及以前并没有受到关注的第 5 外显子受到平衡选择的作用, 在进化中受到了选择压力, 预示着它们在免疫系统的进化过程中可能扮演着重要的角色。

关键词: 人类白细胞抗原; *HLA-Cw* 基因; 基因组全长序列; 克隆和测序分析; 单核苷酸多态性

Single nucleotide polymorphisms in 22 *HLA-Cw* alleles in Chinese Han population

ZENG Jian-Qiang, XU Yun-Ping, WANG Da-Ming, GAO Su-Qing, ZOU Hong-Yan, DENG Zhi-Hui

Shenzhen Key Laboratory of Histocompatibility and Immunogenetics, Shenzhen Blood Center, Shenzhen 518035, China

Abstract: To analyze the molecular genetic polymorphism of full-length *HLA-Cw* gene, a total of 28 samples with known genotypes from Chinese Han population were amplified by long-range PCR using high-fidelity *Pfu* polymerase. A fragment 4.5 kb in length of *HLA-Cw* gene was subjected to cloning and haplotype sequencing. The single nucleotide polymorphisms (SNPs) in all segments of the whole region of *HLA-Cw* gene were analyzed. As a result, we detected 22 different *HLA-Cw* alleles in 28 samples, all of which were submitted to GenBank and the IMGT/HLA Database. Among the 22 *HLA-Cw* alleles, the intronic sequences of *Cw*030301*, *Cw*0706* and *Cw*140201* were firstly elucidated. The novel intronic sequence and the SNPs information may help to design allele-specific primers for accurate sequence-based typing (SBT) and to avoid allele dropout events in SBT test. We aligned all the diploid sequences using ClustalX program and imported them into

收稿日期: 2009-10-13; 修回日期: 2010-01-06

基金项目: 深圳市科技计划项目(编号: 200802116), 广东省科技计划项目(编号: 2008B030301277)和广东省自然科学基金项目(编号: 9451803501004124)资助

作者简介: 曾健强(1957-), 男, 学士, 副主任技师, 研究方向: 输血医学免疫学。Tel: 0755-83366336; E-mail: zjq200411@126.com

通讯作者: 徐筠婷(1982-), 女, 博士, 研究方向: 分子免疫遗传学。Tel: 0755-83242567; E-mail: yunpingxu1982@yahoo.com.cn

Dnasp4.0 to calculate polymorphism in all coding- and non-coding regions. We found 244 SNPs and 10 insertion/deletions (Indels). According to the analysis of polymorphism level, phylogenetic trees and frequency spectrum, we proposed that the evolution of intron 4 and exon 5 was under balancing selection. Selection on these segments indicated that they may be functionally important in evolution of *HLA-Cw* gene. The full-length sequences obtained and related SNPs information can be used as resources of markers for high-resolution typing, complex diseases association studies and human evolution.

Keywords: human leukocyte antigen (HLA); *HLA-Cw* gene; genomic full length; cloning and sequencing analysis; single nucleotide polymorphisms (SNPs)

人类白细胞抗原 Cw (Human leukocyte antigen-Cw, HLA-Cw) 属于经典的 HLA I 类分子。它的功能表现在两个方面, 首先它可结合内源性抗原肽 (Endogenous antigen) 并提呈给 CD8⁺ T 细胞识别, 诱发特异性免疫反应; 另外它还可作为杀伤细胞免疫球蛋白样受体 (Killer immunoglobulin-like receptor, KIR) 的配体, 调节 NK 细胞 (Natural killer cells) 的杀伤功能^[1], 在移植免疫、肿瘤免疫和抗感染免疫中发挥重要作用。

单核苷酸多态性 (Single nucleotide polymorphisms, SNPs) 位点除了分布在 HLA -I、II 类基因的编码抗原结合肽的外显子区, 还大量地分布在 *HLA* 基因的 5 启动子、3 非翻译区以及所有的内含子当中。这些 SNPs 数据蕴含了大量有关基因进化、基因表达调控、疾病相关联的信息, 也是临床和骨髓库进行 HLA 高分辨水平基因分型的分子生物学基础。

2009 年 7 月, 国际 IMGT/HLA 数据库 (<http://www.ebi.ac.uk/imgt/hla/stats.html>) 公布的 *HLA-Cw* 等位基因有 463 个, 在 *HLA-Cw* 等位基因的 DNA 序列中, 提交第 2 和第 3 外显子序列的等位基因较多, 具有完整的 *HLA-Cw* 全长序列的等位基因较少, 特别是缺乏中国人群 *HLA-Cw* 等位基因的全长序列。以往, 我们曾发现因 *Cw*0706* 等位基因非编码区序列和 SNPs 信息的缺乏, 有的商品化测序分型试剂盒检测中国人群样本时存在 *Cw*0706* 等位基因漏检和丢失现象^[2]。本文采用业已建立的 *HLA-Cw* 基因全长序列分子克隆和测序方法^[3], 进行汉族人群 *HLA-Cw* 基因全长序列及其不同亚区 SNPs 多态性分布的分析和研究, 将为中国汉族人群 *HLA-Cw* 基因 SNPs 与疾病相关性的研究以及群体遗传学研究等提供基础。

1 材料和方法

1.1 样本选取及基因组 DNA 的提取

在 620 名已进行 *HLA-Cw* 基因高分辨基因分型、结果已知的汉族造血干细胞捐献者的血样中, 选择 28 份包含汉族个体常见的 *HLA-Cw* 等位基因型的样本用于本研究。这 28 个个体均为常驻深圳的定期无偿献血者, 已签署志愿者知情同意书, 分别来自地域相近的南方五省 (广东 8 例、广西 6 例、湖南 5 例、湖北 5 例、江西 4 例), 互相之间无亲缘关系。每份血样 5 mL, 用 5% EDTA 抗凝。各取 0.25 mL 抗凝血, 采用 E-Z 96TM 全血 DNA 提取试剂盒 (美国 Omega 公司) 提取基因组 DNA。

1.2 *HLA-Cw* 全长序列的分子克隆、测序以及序列拼接

根据文献[3]中的克隆和测序方法, 采用自行设计的一对 *HLA-Cw* 基因全长序列扩增引物, 利用高保真 *Pfu* 酶 (*Pfu* ultra II Fusion HS DNA polymerase, 美国 Stratagene 公司) 特异性地扩增 *HLA-Cw* 基因目的片段 4.5 kb (从上游-962 位碱基至下游 3576 位碱基), 涵盖了 *HLA-Cw* 基因的 8 个外显子和 7 个内含子, 以及上游 5' 调控区和下游 3' 非翻译区。扩增产物克隆至 pGEM-Teasy 载体 (美国 Promega 公司) 后, 提取阳性质粒测序 (质粒小量提取试剂盒购自美国 Axygen 公司)。

测序于 ABI 3730 型测序仪 (美国 ABI 公司) 上进行。将所有测序引物得到的测序峰图文件 (chromas) 输入到 DNASTAR 软件包中的 SeqMan (Lasergene version 5; DNA Star, Inc., Madison, WI) 程序, 按每一样本号分别进行序列自动拼接。人工核查排除 PCR 突变, 得到单个样品最终的两条单倍体序列。再将

所有样品的序列输入 Clustal-X 软件^[4]排比, 输出一个群体比对数据。排比后的文件再放入 Bioedit v6.0.7 软件中进行最后的人工核查。

1.3 群体遗传学统计分析

1.3.1 基因多态性分析

π 和 θ_w 通过 DnaSP 4.0^[5] 计算。 π 是 Tajima 估计值^[6], 为每个位点上序列两两比较后的平均核苷酸多样性, π 值大小主要受中高频突变影响。 θ_w 是 Watterson 估计值^[7], 为每个位点上差异核苷酸的数目, 计算公式为: $\theta_w = S/a_n$ ($a_n = 1 + 1/2 + \dots + 1/(n-1)$)。S (Segregating Sites) 即分离位点数, 并不考虑突变碱基的频率。非同义突变率 (dn) 及同义突变率 (ds) 用 DnaSP 4.0 软件计算^[8]。在中性选择下, 两者的比例应该相等。若非同义突变率大于同义突变率则提示可能有正选择存在; 反之则提示可能存在负选择^[9]。

1.3.2 SNPs 的频谱分析 (Frequency spectrum)

多态位点 P_i ^[10] 的频谱定义为: 在一个含有 n 个个体的群体中, 出现 i 次的多态位点的数目。中性模型的 P_i 即其期望值通过公式计算:

$$P_i = \frac{\theta_w}{k} = \frac{1}{a_n} * \frac{1}{i}$$

$$a_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n-1}$$

其中, k 是多态位点的总个数。观察值 P_i 通过直接计数获得。本文中提到的 Minor allele (频率 < 50% 的突变碱基) 频谱图中的期望值按 $\tilde{P}_i = P_i + P_{n-i} \quad i: i \leq \frac{n}{2}$ 来计算。将观察值与中性状态下的期望值相比, 根据其偏离情况来判断其是否受自然选择。一般来说, 如果是正选择或负选择, 低频率突变碱基 (< 10%) 的个数将远大于期望值。如果受到平衡选择, 中等频率突变碱基 (20%~50%) 的个数将远大于期望值。

1.3.3 系统发育树分析

HLA-Cw 基因亚区的系统发育树使用 MEGA3.1^[11] 中的 Neighbor-Joining 方法构建。其中, 距离矩阵通过 Kimura 双参数法 (Kimura's two-parameter method) 计算。树的可靠性由 Bootstrapping 重抽样 1 000 次来估计。另外, 不同灵长类的系统发育分析, 可以比

较种内差异及种间差异的程度, 从而判断区域是否存在跨物种多态性。黑猩猩 (*Pan troglodytes*)、猩猩 (*Pongo satyrus*) 的 *HLA-Cw* 序列应用 Blat 程序从 UCSC 基因组数据库 (<http://genome.ucsc.edu>) 中获得, 用于跨物种多态性的比较。

1.3.4 连锁不平衡 (Linkage disequilibrium, LD) 分析

为判断 *HLA-Cw* 基因全长序列中各亚区 SNPs 之间的连锁与重组情况, 我们通过绘制连锁不平衡图谱来进行分析。连锁不平衡主要使用 Windows Haploview version 4.0 (<http://www.broad.mit.edu/mpg/haploview/>) 分析。根据 D' 值^[12] 的大小与统计学检验是否显著, 以不同颜色绘制连锁不平衡图谱。 D' 值等于 1 表示 SNPs 之间完全连锁 (没有重组事件发生过), 用亮红色或蓝色方块表示, 其中有统计学意义的 ($LOD > 2$) 是亮红色, 无统计学意义的用蓝色表示。浅红色表示 D' 值小于 1 但有统计学意义, 意味着 SNPs 之间有较强的连锁关系。白色表示 D' 值较小甚至接近于 0, 意味着 SNPs 之间重组较频繁。

2 结果与分析

2.1 22 条 *HLA-Cw* 等位基因全长序列以及与 IMGT/HLA 数据库的比较

通过测序、拼接、排比、筛查, 最终获得了 28 个样本的 56 条 *HLA-Cw* 等位基因的全长序列 (-962 bp~3 576 bp)。56 条序列中共发现 22 种不同的单倍型, 这些序列已提交给 GenBank 获得登录号 (表 1), 同时向 IMGT/HLA 数据库提交并获得世界卫生组织 (World Health Organization, WHO) 命名委员会的命名。在 22 条单倍型中, 共存在 244 个单核苷酸多态性, 10 个插入/缺失多态性, 其中 $Cw^*04010103$ 是我们发现的一个新等位基因, 它与 $Cw^*04010101$ 的区别在于第 3 内含子有一个碱基的变异 (nt 1111 G > A), 此外, 在中国人群中发现的 $Cw^*030401$ 等位基因在 3'-UTR 区域的 nt 3033 位均为 "C", 而 IMGT/HLA 数据库中所公布的 $Cw^*030401$ 的序列在该碱基位置为 "G", 显示中国个体 $Cw^*030401$ 的特殊性。

2.2 *HLA-Cw* 各区域多态性水平比较

将 *HLA-Cw* 基因 56 条全长序列划分为 16 个亚区, 计算各亚区的平均核苷酸突变率 (π 与 θ_w)。此外,

表 1 HLA-Cw 各等位基因 GenBank 登录号及 IMGT/HLA 数据库注册号

等位基因名称	GenBank 登录号	IMGT/HLA 注册号	长度(bp)	等位基因名称	GenBank 登录号	IMGT/HLA 注册号	长度(bp)
*010201	FJ515900	HWS10006562	4 537	*070401	FJ785728	HWS10006550	4 551
*010201v1	FJ827032	HWS10006744	4 537	*0706	FJ785732	HWS10006238	4 551
*030202	FJ827033	HWS10006563	4 536	*0766	FJ785729	HWS10006551	4 551
*030301	FJ515902	HWS10006152	4 537	*0767	FJ785730	HWS10006552	4 551
*030401	FJ515903	HWS10005879	4 537	*080101	FJ785724	HWS10006554	4 536
*04010101	FJ515901	HWS10006561	4 537	*0802	FJ785733	HWS10006555	4 536
*04010103	FJ515899	HWS10005875	4 537	*0820	FJ785725	HWS10005896	4 536
*06020101	FJ785734	HWS10006557	4 536	*0821	FJ785726	HWS10006556	4 536
*06020102	FJ785735	HWS10006558	4 536	*0822	FJ785727	HWS10006188	4 536
*07020101	FJ515904	HWS10006548	4 551	*140201	FJ785736	HWS10006239	4 537
*07020103	FJ785731	HWS10006549	4 551	*150201	FJ785737	HWS10006559	4 537

表 2 HLA-Cw 基因各区域多态性水平

HLA-Cw 各亚区	长度(bp)	S	$\pi (\times 1000)$	$\theta_w (\times 1000)$	$D_{hc} (\times 100)$	$ds (\times 1000)$	$dn (\times 1000)$
5 -启动子	970	28	8.98	6.36	1.10	—	—
第 1 外显子	73	4	20.76	11.93	1.91	0	21.85
第 1 内含子	130	10	19.91	16.75	4.34	—	—
第 2 外显子	270	21	24.75	16.93	2.75	31.95	21.79
第 2 内含子	251	13	20.20	11.55	2.86	—	—
第 3 外显子	276	27	31.77	21.30	3.88	30.17	32.75
第 3 内含子	587	30	13.63	11.13	1.71	—	—
第 4 外显子	276	17	19.54	13.41	1.70	36.87	14.09
第 4 内含子	124	13	34.76	23.39	3.17	—	—
第 5 外显子	120	13	33.88	23.58	3.65	43.46	27.83
第 5 内含子	441	19	14.26	9.44	1.40	—	—
第 6、7、8 外显子	86	4	9.20	10.13	1.72	0	12.72
第 6 内含子	107	4	10.58	8.14	1.18	—	—
第 7 内含子	164	8	15.43	10.62	2.23	—	—
3 -非翻译区	416	23	19.59	12.09	2.42	—	—
3 -下游区域	267	10	13.17	8.18	2.41	—	—

注: S: 分离位点数; π : 序列两两比较平均核苷酸突变率; θ_w : 平均每核苷酸分离位点数; ds: 非同义突变率; dn: 同义突变率; Dhc: 人与黑猩猩种间变异率; 表中阴影标注的是多态性较高且在文章重点说明的区域。

利用 UCSC 网站(<http://www.genome.ucsc.edu>)上的 Blat 程序获得了黑猩猩的 Cw 基因全长序列计算各亚区的种间差异(D_{hc})。如表 2 所示, 编码抗原结合肽的外显子(第 2、3 外显子)的种内变异率与种间差异均显示较高的水平, 且在第 3 外显子中非同义突变率(32.75/kb)明显大于同义突变率(30.17/kb)。

另外, HLA-Cw 大部分内含子的平均核苷酸突变率($10 \sim 20$ /kb)远比人类基因组平均核苷酸突变率 0.75/kb^[13]高, 然而与外显子的同义突变率(30/kb)相比又显著下降。我们分析了 HLA-Cw 基因的连锁不平衡图。由图 1 所示, HLA-Cw 基因第 2、3 外显

子与其周边的第 1、2 及第 3 内含子有重组事件发生(白色方块), 这些重组事件会减弱外显子搭车效应的影响, 因而解释了内含子较之外显子同义突变率变低的现象。

除了第 2、3 外显子突变率较高外, 第 4、5 外显子的突变率水平也较高。尤其是第 5 外显子, 它的同义突变率达到 43.46/kb。与第 5 外显子临近的第 4 内含子的平均核苷酸突变率(π)也是所有区域中最高的(34.76/ kb)。值得一提的是, 第 5 外显子与第 4 内含子的多态性模式非常相似, 无论是平均核苷酸突变率(π 与 θ_w), 还是种间核苷酸突变率(D_{hc}), 它

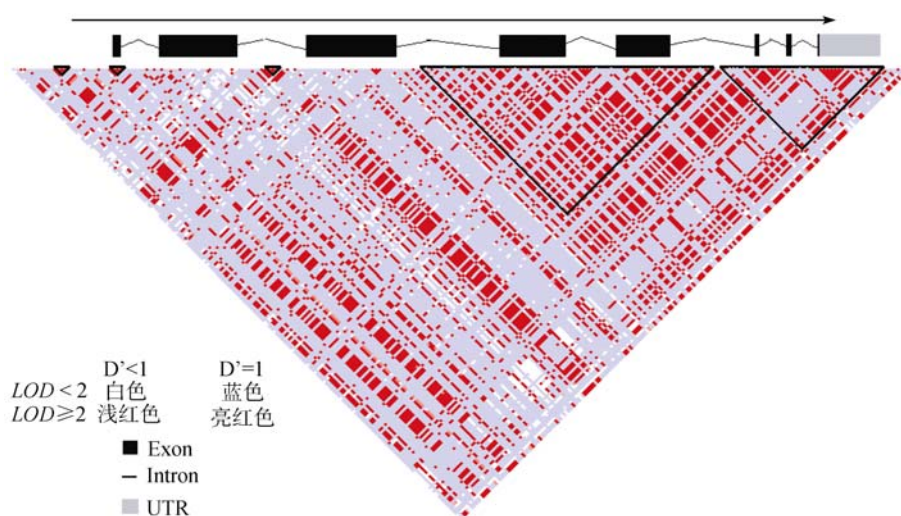


图 1 *HLA-Cw* 全基因连锁不平衡图

SNPs 两两之间的连锁关系用矩阵作图, D' 值的大小及统计学意义用不同颜色方块显示(见材料和方法 1.3.4)。

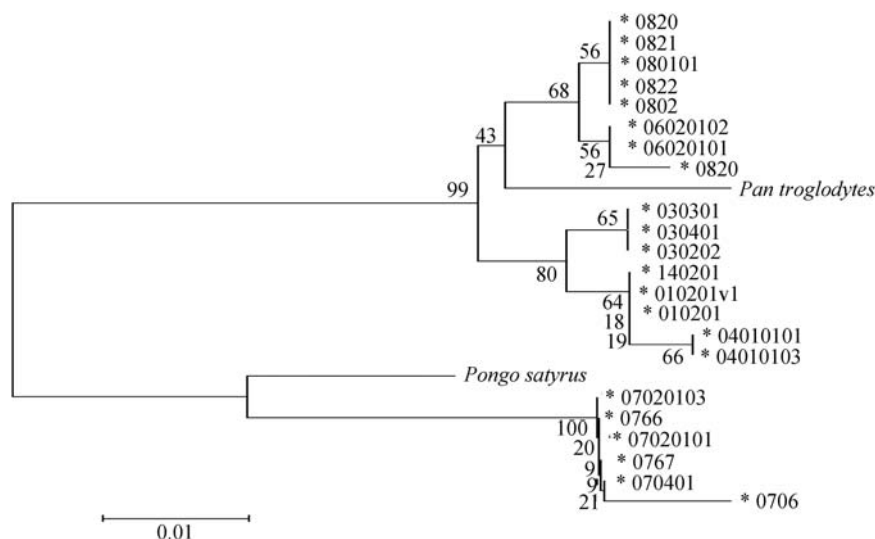


图 2 第 4 内含子及第 5 外显子系统发育树

人类序列以等位基因名称(如表 1)注明。

们的值在两个区域中非常接近(表 2); 并且从图 1 中我们可以看出第 4 内含子及第 5 外显子中处于同一个紧密连锁的单倍型域(所有 SNPs 之间的 D' 值 > 0.9 , 表示极少有重组事件发生)。

2.3 第 4 内含子及第 5 外显子进化树与频谱图分析

为了探讨 *HLA-Cw* 基因的第 4 内含子及第 5 外显子在进化过程中是否受到选择压力, 本文利用 UCSC 网站(<http://www.genome.ucsc.edu>)上的 Blat 程序获得了与人类距离更远的猩猩的序列, 并与黑猩猩的序列一起来对两个区域作进化树分析。人类的序列分为两支, 而黑猩猩与猩猩的序列分散在人类

序列的两个分支当中, 见图 2。Cw*07 家系作为独立的一支与猩猩的序列簇立在一起, 说明这一家系作为古老单倍型已保存了约 1 400 万年之久(猩猩分化于 1 400 万年前)。这种跨物种多态性揭示了 *HLA-Cw* 基因的第 4 内含子与第 5 外显子在进化过程中受到了平衡选择的作用。

此外, 我们对第 4 内含子及第 5 外显子区域的多态位点作了频谱分析。由于跨物种多态性的存在, 我们无法定义祖先等位基因, 因此我们绘制了此区域内小频率等位基因(Minor allele, 频率小于 50% 的突变碱基)的频谱分布图。若受平衡选择作用, 频谱

图内近中等频率的 SNPs 会远远超过中性期望值。由图 3 可知, 个数最多的是频率为 $(12/56 = 21.4\%)$ 的 SNPs, 频率为 26.8% $(15/56)$ 以及 48.2% $(27/56)$ 的 SNPs 的个数都高于期望值, 说明第 4 内含子与第 5 外显子符合平衡选择的模式。

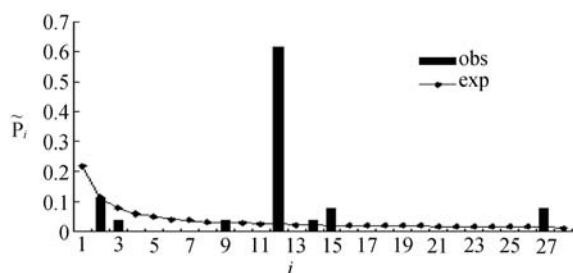


图 3 *HLA-Cw* 第 4 内含子及第 5 外显子频谱图

obs 表示观察值, exp 表示期望值。

3 讨论

当前, *HLA* 经典基因全长序列多态性的研究正悄然兴起。这是基于两方面研究的需求, 一是基因的免疫调控机制研究的需要。众所周知, 以往对 *HLA* 多态性的研究主要是针对编码抗原结合肽的外显子(第 2、3 外显子)。由于功能的重要性, 众多科学家对这些外显子的多态位点的探测、疾病相关性以及进化分析研究做了大量工作; 到上世纪 90 年代中, 科学家们开始投入到对 *HLA* 基因的非编码区特别是 5'-启动子的多态性研究当中, 并从进化、功能分析等多种角度清晰地揭示了这些位于非编码区中的 SNPs 在基因表达调控中的作用^[14~16]。

二是临床高分辨配型的需要。截止到 2009 年 4 月, IMGT/HLA 数据库公布的 *HLA-Cw* 位点的等位基因数目为 440 个, 但具有全长序列的等位基因为 49 个, 大量等位基因的全长序列有待填补。*HLA-Cw* 基因全长序列及 SNPs 信息, 为 *HLA-Cw* 基因高分辨分型试剂开发时设计引物和探针的重要依据。在本文中, 我们共测定 22 种等位基因, 其中 *Cw*0706*、*Cw*030301*、*Cw*140201* 的全长序列为首次报道, 这 3 种等位基因的全长序列已向 IMGT/HLA 数据库提交, 目前 *Cw*0706*、*Cw*030301* 及 *Cw*140201* 已被正式收录到该数据库中的“Alignments”模块中。

与 IMGT/HLA 数据库中所公布的 *HLA-Cw* 基因序列相比, 本文中所检测到的 *HLA-Cw* 全长序列有

以下几个特点: 第一, IMGT/HLA 数据库中公布的 *HLA-Cw* 全长序列位置从 nt-283 ~nt3066, 而本文中的 *HLA-Cw* 全长序列 (nt-962~nt3358) 延伸了 IMGT/HLA 数据库所涵盖的区域, 共发现了 244 个单核苷酸多态性, 10 个插入/缺失多态性。在 IMGT/HLA 没有覆盖到的启动子区域(nt-962~nt-284), 我们共发现了 17 个 SNPs 以及 3 个插入缺失多态性, 这些 SNPs 形成了 11 种单倍型, 其中 *Cw*07* 家系的单倍型与其他家系的单倍型差别最大。Huard 等^[17]的研究表明: *Cw*07* 的表达量为其他 *HLA-Cw* 等位基因的两倍, 那么位于 *Cw*07* 基因 5'-启动子区上游的 SNPs 单倍型是否对这个结果有直接的影响, 有待功能实验的验证。第二, 在 IMGT/HLA 数据库中, 等位基因 *Cw*140201*、*Cw*0706*、*Cw*030301* 只有编码区的 8 个外显子序列, 内含子等非编码区序列为我们首次报道。尤其是 *Cw*0706* 内含子序列的获得, 使得我们能够重新设计对该等位基因测序分型的引物, 解决商品化测序分型试剂盒对这一等位基因漏检的问题^[2]。第三, *Cw*04010103* 是我们发现的一个新等位基因, 它与 *Cw*04010101* 的区别在于第三内含子的 nt 1111 位有一个碱基的变异(*Cw*04010101G-Cw*04010103A*)。第四, 在中国人群中发现的 *Cw*030401* 等位基因在 3'-UTR 区域的 nt 3033 位均为“C”, 而 IMGT/HLA 数据库中所释放的 *Cw*030401* 的序列在 nt 3033 位为“G”, 显示中国个体 *Cw*030401* 的特殊性。这些获得的序列, 均向 IMGT/HLA 数据库提交并被该数据库收录, 丰富了 *HLA-Cw* 基因的多态性数据。

我们将 *HLA-Cw* 基因 56 条全长序列划分为 16 个亚区, 利用群体遗传学方法分析比较各个亚区的多态性水平。我们发现第 2、3 外显子突变率较高, 且引起氨基酸改变的非同义突变率大于同义突变率, 说明了这些编码区受到强烈的平衡选择作用, 与 Hughes 等^[14]的研究结果相符。另外, *HLA-Cw* 大部分内含子的平均核苷酸突变率远比人类基因组平均核苷酸突变率高, 说明它们受到了外显子平衡选择的搭车效应(hitchhiking)的影响。然而内含子的变异率又显著低于外显子的同义突变率, 对此 Cereb 等^[15]也提出了合理的解释: 外显子的同义突变位点由于与受选择的非同义突变位点之间由于甚少重组紧密连锁而显示高度的多态性, 而内含子与外显子间有

可能发生频繁重组而产生相对较低的多态性。因此, 根据连锁不平衡分析的结果以及 Cereb 等人提出的理论, 我们判断 *HLA-Cw* 基因内含子所表现出来的多态性特点应该是外显子的搭车效应与重组双重作用的结果, 它们本身并没有受到选择。

本文中值得注意的是 *HLA-Cw* 基因的第 4、5 外显子的突变率水平也较高。在本文中, 我们利用连锁不平衡分析, 发现第 4 内含子及第 5 外显子紧密连锁; 对该区域进行进化树分析, 揭示了 *HLA-Cw* 基因的第 4 内含子与第 5 外显子在进化过程中受到了平衡选择的作用, 并且通过频谱图分析证明了第 4 内含子与第 5 外显子符合平衡选择的模式。由于第 4 内含子与第 5 外显子紧密连锁, 它们之间协同进化, 哪个区域在免疫系统中扮演更重要的作用并不清楚, 有以下几种可能性: 第一, 第 5 外显子编码的是跨膜区和胞浆区尾部的初始部分, 它的多态性在个体免疫应答多样性中可能与第 2、3 外显子中一样, 扮演着重要的作用, 如最近的研究显示, *HLA-I* 类基因跨膜区可以结合某些细胞因子来调控基因的表达^[18], 而第 4 内含子由于与第 5 外显子完全连锁, 它的多态性没有任何功能意义, 只是单纯受到它的搭车效应。第二, 不仅第 5 外显子的多态性有功能, 第 4 内含子的多态性在基因表达调控中同样发挥重要作用, 因此二者协同进化, 共同发挥功能。

参考文献(References):

- [1] Colonna M, Brooks EG, Falco M, Ferrara GB, Strominger JL. Generation of allospecific natural killer cells by stimulation across a polymorphism of *HLA-C*. *Science*, 1993, 260 (5111): 1121–1124.
- [2] 曾健强, 徐筠娉, 王大明, 邹红岩, 邓志辉, 杨宝成. 5 例样本 *HLA-Cw* 基因测序分型结果中等位基因丢失及其原因分析. *中华医学遗传学杂志*, 2009, 26(5): 562–566.
- [3] 邓志辉, 徐筠娉, 高素青, 李大成, 喻琼, 苏宇清, 曾健强, 杨宝成. *HLA-Cw* 基因全长序列分子克隆及测序方法的建立. *中华医学遗传学杂志*, 2009, 26(3): 258–262.
- [4] Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. *Nucleic Acids Res*, 1997, 25(24): 4876–4882.
- [5] Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 2003, 19(18): 2496–2497.
- [6] Tajima F. Statistical analysis of DNA polymorphism. *Jpn J Genet*, 1993, 68(6): 567–595.
- [7] Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*, 1975, 7(2): 256–276.
- [8] 陈玲玲, 彭贵子, 张伟丽, 田大成. 突变在基因组进化中的意义. *遗传*, 2006, 28(5): 631–638.
- [9] Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet*, 2002, 18(9): 486.
- [10] Fu YX. Estimating effective population size or mutation rate using the frequencies of mutations of various classes in a sample of DNA sequences. *Genetics*, 1994, 138 (4): 1375–1386.
- [11] Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinformatics*, 2004, 5(2): 150–163.
- [12] Lewontin RC. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, 1964, 49(1): 49–67.
- [13] Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter DJ, Lander ES, Altshuler D. Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet*, 2002, 32(1): 135–142.
- [14] Hughes AL, Yeager M. Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet*, 1998, 32: 415–435.
- [15] Cereb N, Hughes AL, Yang SY. Locus-specific conservation of the *HLA* class I introns by intra-locus homogenization. *Immunogenetics*, 1997, 47(1): 30–36.
- [16] Andersen LC, Beaty JS, Nettles JW, Seyfried CE, Nepom GT, Nepom BS. Allelic polymorphism in transcriptional regulatory regions of *HLA-DQB* genes. *J Exp Med*, 1991, 173(1): 181–192.
- [17] Huard B, Karlsson L. KIR expression on self-reactive CD8+ T cells is controlled by T-cell receptor engagement. *Nature*, 2000, 403(6767): 325–328.
- [18] Brémont A, Meynet O, Mahiddine K, Coito S, Tichet M, Scotlandi K, Breittmayer JP, Gounon P, Gleeson PA, Bernard A, Bernard G. Regulation of *HLA* class I surface expression requires CD99 and p230/golgin-245 interaction. *Blood*, 2009, 113(2): 347–357.